Sampling through Random Walks*

H. Edwin Romeijn

Department of Operations Research & Tinbergen Institute

Erasmus University Rotterdam

P.O. Box 1738

3000 DR Rotterdam, The Netherlands

Robert L. Smith

Department of Industrial and Operations Engineering

The University of Michigan

Ann Arbor, Michigan 48109-2117

August 1989
Revised December 1989

---

1 Introduction

In this paper we will discuss the following problem: Given a bounded set $S \subset \mathcal{R}^n$, how can we efficiently sample points $X_1, X_2, \ldots$ which are distributed according to a probability density function f over S. That is, we require that the probability that $X \in A \subseteq S$ be given by $\int_A f(x)\, dx =: F(A)$.

In [Smith, 1984] a class of symmetric mixing algorithms (later called "hit-and-run algorithms", see [Berbee et al., 1987]) was introduced, which generate a sequence of points in a bounded, open region S, having the property that the limiting distribution of this sequence is the uniform distribution, independent of the initial point.

In [Boender et al., 1989] a class of "shake-and-bake" algorithms is discussed, which generate a sequence of points on the boundary of a polytope with a uniform limiting distribution. Recently this class of algorithms has been extended to the boundary of arbitrary bounded convex bodies [McDonald, 1989].

We will generalize both classes of algorithms mentioned above to obtain a limiting distribution given by F. Conditions on F (or f) will be given for this to be possible. Possible applications of these algorithms include the estimation of volumes or surface areas. For example, the surface area of engine components is important to General Motors, for estimating heat loss over time. Also, these methods can be used for Monte Carlo integration, which is important in Bayesian statistics when the dimension of the problem is too high for numerical integration to be efficient. A third possible application is in global optimization. We will study this application in this paper.

The outline of the paper is as follows. In section 2 we will introduce a general class of Random Walks, and give some examples. Also, we will give conditions under which a Random Walk has a limiting distribution given by a density function f. In section 3 we will define a class of "hide-and-seek" algorithms for global optimization.

2 Random Walks

2.1 The general class

Consider bounded set $S \subset \mathcal{R}^n$. The general random walk can be described as follows.

Step 0: Choose a starting point $x^0 \in S$, and set $i := 0$.
Step 1: Generate a direction d^i from an arbitrary probability distribution over a direction

1

set D. Find the set $\Lambda^i = \{\lambda \in \mathcal{R} | x^i + \lambda d^i \in S\}$.

Step 2: Generate a real number λ^i from an arbitrary probability distribution over the set Λ^i.

Step 3: Set $x^{i+1} := x^i + \lambda^i d^i$ and set $i := i + 1$. Go to *Step 1*.

Clearly, the random sequence of points $\{X^i\}_{i=1}^{\infty}$ constitute a continuous state homogeneous Markov chain. We assume that the transition density function $p(y|x)$ corresponding to this chain exists. Furthermore, we assume that there exists a continuous density function f over S such that

$$p(y|x)f(x) = p(x|y)f(y) \qquad \forall\; x, y \in S.$$

Algorithms satisfying these assumptions will be called Random Walks. An example is the hypersphere directions (HD) hit-and-run algorithm, discussed in [Boneh & Golan, 1979], [Smith, 1980], [Smith, 1984] and [Berbee et al., 1987]. In this algorithm the distributions in steps 1 and 2 are chosen to be the uniform distribution, and the direction set D is the unit hypersphere. The density f is the uniform density over S.

2.2 Examples

In the following we will assume that a continuous density function f on S is given.

For *Random Walk 1* the region S must be open. Furthermore, the direction set D is the unit hypersphere, and a direction will be drawn from the uniform distribution over D. We choose the density of $\underline{\lambda}^i$ to be the marginalization of f on the line set defined by Λ^i:

$$f_i(\lambda) = \frac{f(x^i + \lambda d^i)}{\int_{\Lambda^i} f(x^i + \mu d^i)\, d\mu} \qquad \text{for } \lambda \in \Lambda^i.$$

Theorem 2.1 *The transition density function corresponding to* Random Walk 1 *reads:*

$$p_1(y|x) = \frac{2}{C_n \|x - y\|^{n-1}} \cdot \frac{f(y)}{\int_{\Lambda(x,y)} f\left(x + \lambda \frac{y-x}{\|y-x\|}\right) d\lambda} \qquad \textit{for all } x \neq y \in S,$$

where C_n *is the surface area of the n-dimensional unit hypersphere, and* $\Lambda(x,y) = \left\{\lambda \in \mathcal{R} | x + \lambda \frac{y-x}{\|y-x\|} \in S\right\}$.

Proof

For all $y \in S$, $y \neq x$ the transition density function can be defined as:

$$p_1(y|x) = \lim_{\varepsilon \downarrow 0} \frac{\Pr\left\{X^1 \in H_\varepsilon(y) | X^0 = x\right\}}{m_n(H_\varepsilon(y))}$$

2

where $H_\varepsilon(y)$ is an n-dimensional cube of content ε^n centered at y and oriented along the ray from x to y, and where $m_n(\cdot)$ denotes the n-dimensional Lebesgue measure. We have:

$$
\begin{aligned}
p_1(y|x) &= \lim_{\varepsilon \downarrow 0} \frac{\Pr\left\{X^1 \in H_\varepsilon(y) | X^0 = x\right\}}{m_n(H_\varepsilon(y))} \\
&= \lim_{\varepsilon \downarrow 0} \frac{2\varepsilon^{n-1}}{C_n \|x - y\|^{n-1}} \cdot \frac{\varepsilon\, f(y)}{\int_{\Lambda(x,y)} f\left(x + \lambda \frac{y-x}{\|y-x\|}\right) d\lambda} \cdot \frac{1}{\varepsilon^n} \\
&= \frac{2}{C_n \|x - y\|^{n-1}} \cdot \frac{f(y)}{\int_{\Lambda(x,y)} f\left(x + \lambda \frac{y-x}{\|y-x\|}\right) d\lambda}
\end{aligned}
$$

which proves the theorem. ∎

In *Random Walk 2* the class of regions S, the direction set D and the distribution of the direction will be the same as for *Random Walk 1*. The density of $\underline{\lambda}^i$ is a mixed discrete/continuous one:

$$
\Pr\{\underline{\lambda}^i = 0\} = 1 - \beta(y|x) \qquad \text{and}
$$

$$
f_i(\lambda) = \frac{\beta(y|x)}{d(x,y)} \qquad \text{for } \lambda \in \Lambda^i,
$$

where $d(x,y)$ is the diameter of S along the line through x and y, and $\beta(y|x)$ is the so-called move probability function (see [Boender et al., 1989]), which satisfies the following conditions:

$$
0 \le \beta(y|x) \le 1 \qquad \text{for all } x, y \in S
$$

$$
\beta(y|x)\, f(x) = \beta(x|y)\, f(y) \qquad \text{for all } x, y \in S.
$$

Theorem 2.2 *The transition density function corresponding to* Random Walk 2 *reads:*

$$
p_2(y|x) = \frac{2}{C_n \|x - y\|^{n-1}} \cdot \frac{\beta(y|x)}{d(x,y)}
$$

for $x \ne y \in S$, and

$$
p_2(x|x) := \Pr\{X^1 = x | X^0 = x\} = 1 - \int_S p_2(y|x)\, dy
$$

for all $x \in S$.

Proof
Analogous to the proof of theorem 2.1. ∎

In *Random Walk 3* S will be the boundary of a full-dimensional polytope. The direction set D is, once again, equal to the unit hypersphere. The density of the direction will be expressed in terms of the "hitpoint" y, which is the intersection of S with the line through a point $x \in S$ with a certain direction vector, and denoted by $\rho(y|x)$. We assume that this density is absolutely continuous. The density of $\underline{\lambda}^i$ is again a mixed discrete/continuous one:

$$\Pr\{\underline{\lambda}^i = 0\} = 1 - \beta(y|x) \qquad \text{and}$$

$$\Pr\{\underline{\lambda}^i = \|y - x\|\} = \beta(y|x).$$

Note that in this case $\Lambda^i = \Lambda(x,y) = \{0, \|y - x\|\}$. Furthermore, $\beta(y|x)$ is the move probability function, satisfying the same conditions as in *Random Walk 2*.

Theorem 2.3 *The transition density function corresponding to* Random Walk 3 *reads:*

$$p_3(y|x) = \beta(y|x)\,\rho(y|x)$$

for $x \neq y \in S$, and

$$p_3(x|x) := \Pr\{X^1 = x | X^0 = x\} = 1 - \int_S p_3(y|x)\,dy$$

for all $x \in S$.

Proof
See [Boender et al., 1988] and [Boender et al., 1989]. ∎

2.3 Limiting density

Lemma 2.4 *If the transition density function p of a Random Walk on a set S satisfies*

$$p(y|x)\,f(x) = p(x|y)\,f(y) \qquad \text{for all } x \neq y \in S$$

where f is a continuous density on S, then f is a stationary density of the Markov chain $\{X^i\}_{i=1}^{\infty}$, i.e.

$$\int_S \Pr\{X^{i+1} \in A | X^i = x\}\,f(x)\,dx = F(A)$$

for all $A \subseteq S$.

4

Proof

$$\int_S \Pr\{X^{i+1} \in A | X^i = x\} f(x)\, dx =$$

$$= \int_S \int_A p(y|x) f(x)\, dy\, dx + \int_S \Pr\{X^{i+1} = x | X^i = x\}\, 1_A(x)\, f(x)\, dx$$

$$= \int_S \int_A p(x|y) f(y)\, dy\, dx + \int_A \left\{ 1 - \int_S p(y|x)\, dy \right\} f(x)\, dx$$

$$= \int_A \int_S p(x|y) f(y)\, dx\, dy + \int_A f(x)\, dx - \int_A \int_S p(y|x) f(x)\, dy\, dx$$

$$= \int_A f(x)\, dx = F(A). \quad \blacksquare$$

Theorem 2.5 *If f is a stationary density of a Random Walk, and the absolutely continuous component of the transition density function is uniformly bounded away from zero, then f is the limiting density of the Markov chain generated by the Random Walk.*

Proof
Since the absolutely continuous component of the transition density function p of any Random Walk is uniformly bounded away from zero, the stationary density f is unique, and the limiting density of the corresponding Markov chain is equal to f (see [Doob, 1953]). $\blacksquare$

2.4   Examples (continued)

In this section we will derive more specific conditions on the Random Walks described in section 2.1 such that these algorithms have a limiting density given by f.

Theorem 2.6 *If*

$$f_{min} := \inf_{x \in S} f(x) > 0$$

$$f_{max} := \sup_{x \in S} f(x) < \infty$$

then the limiting density of the Markov chain generated by Random Walk 1 is f.

Proof
Under the conditions mentioned in the theorem the transition density function p_1 is uniformly

bounded away from zero:

$$p_1(y|x) = \frac{2}{C_n\,\|x-y\|^{n-1}} \cdot \frac{f(y)}{\int_{\Lambda(x,y)} f(x + \lambda \frac{y-x}{\|y-x\|})\,d\lambda}$$

$$\geq \frac{2}{C_n\,r_S^{n-1}} \cdot \frac{f_{min}}{r_S\,f_{max}} > 0$$

where $r_S < \infty$ is the maximum diameter of S:

$$r_S = \max_{x,y \in S} \|x - y\|.$$

Now use Theorem 2.5. ∎

Theorem 2.7 *Suppose the conditions in Theorem 2.6 hold. Possible choices for the move probability function β, such that* Random Walk 2 *generates a sequence of points having a limiting density given by f are:*

$$\beta(y|x) = \frac{f(y)}{f_{max}}$$

$$\beta(y|x) = \frac{f_{min}}{f(x)}$$

$$\beta(y|x) = \frac{f(y)}{f(x) + f(y)}$$

$$\beta(y|x) = \min\{1, f(y)/f(x)\}.$$

Proof

The absolutely continuous component of the transition density function p_2 is uniformly bounded away from zero:

$$p_2(y|x) = \frac{2}{C_n\,\|x-y\|^{n-1}} \cdot \frac{\beta(y|x)}{d(x,y)}$$

$$\geq \frac{2}{C_n\,r_S^{n-1}} \cdot \frac{f_{min}/2f_{max}}{r_S} > 0.$$

Now use Theorem 2.5. ∎

Consider the HD algorithm (see section 2.1), and choose S to be a strip of width ε around the boundary of a polytope. Letting ε go to zero gives as a "natural" choice for the acceptance probability function β of *Random Walk 3*:

$$\beta(y|x) = \frac{1/\cos\phi_y}{1/\cos\phi_x + 1/\cos\phi_y}.$$

It is possible to prove that the algorithm *Random Walk 3* with this acceptance probability function generates a uniform limiting distribution on the boundary of a polytope (see [Boender et al., 1989]) if the direction vector is uniformly distributed over the unit hypersphere (see [Boender et al., 1989]) . If, instead of a strip of width ε, we choose S to be a strip of width $\varepsilon f(x)$ around the boundary of a polytope, we get the following choice for the acceptance probability function by letting ε go to zero again:

$$\beta_f(y|x) = \frac{f(y)/\cos\phi_y}{f(x)/\cos\phi_x + f(y)/\cos\phi_y}.$$

Theorem 2.8 *Suppose the conditions in Theorem 2.6 hold. Then we have that the* Random Walk 3 *algorithm with the direction vector uniformly distributed on the unit hypersphere and acceptance probability function β_f generates a limiting density on S given by f.*

Proof

Using the result from [Boender et al., 1989] for the case where f is the uniform density, and using Theorem 2.5 the result follows immediately. ∎

In [Boender et al., 1989] a class of so-called "SB algorithms" is discussed. These algorithms generate a sequence of points on the boundary of a full-dimensional polytope having a uniform limiting distribution. In the following theorem we will consider an element of this class of algorithms, characterized by the density of the "hitpoints" $\rho(y|x)$ and the move probability function $\beta(y|x)$, like in *Random Walk 3*. We will modify the move probability function in such a way that the resulting version of *Random Walk 3* will have a limiting density given by f.

Theorem 2.9 *Suppose the conditions in Theorem 2.6 hold. Furthermore, suppose an SB algorithm is characterized by the density of the hitpoints ρ and the move probability function β. Then the* Random Walk 3 *algorithm characterized by the same hitpoint-density, and by a move probability function given by $\gamma(f(y)|f(x), f_{min}, f_{max}) \cdot \beta(y|x)$, such that:*

$$0 < \delta_\gamma \le \gamma(f(y)|f(x), f_{min}, f_{max}) \le 1 \qquad \text{for all } x, y \in S$$

$$\gamma(f(y)|f(x), f_{min}, f_{max}) f(x) = \gamma(f(x)|f(y), f_{min}, f_{max}) f(y) \qquad \text{for all } x, y \in S.$$

has a limiting density given by f.

Proof

Using the result from [Boender et al., 1989] for the case where f is the uniform density, and using Theorem 2.5 the result follows immediately. ∎

3 Hide-and-seek algorithms

Consider the following mathematical programming problem:

$$\max_{x \in S} f(x)$$

where

1. $S \subset \mathcal{R}^n$ is full dimensional and compact.

2. f is twice continuously differentiable.

It is well-known that the number of iterations necessary to get within ε of the optimal solution increases exponentially in the dimension n of the problem if we use Pure Random Search to approximate the optimal solution of this problem (see [Dixon & Szegö, 1978?], [Rinnooy Kan & Timmer, 1984]).

In this section we will discuss algorithms which generate points from a density function which is a positive, nondecreasing transformation of the objective function f. We will call these *hide-and-seek algorithms*.

Using the methods discussed in this paper we can generate points in the interior of S which are asymptotically distributed to some density function defined on S. [Rubinstein, 1981] suggests that points should be generated from the density

$$g(x; \lambda) \propto e^{\lambda f(x)} \tag{1}$$

for large values of λ. [Pincus, 1968] showed that the expected value of a random variable defined on S with this density converges (for $\lambda \to \infty$) to the point at which the global maximum of f on S is attained, under the condition that this global maximum is unique. We conjecture that the distribution of this random variable for $\lambda \to \infty$ converges to a (degenerate) distribution which is uniform over the set of global maxima. Note also that simulated annealing algorithms, which have mainly been used for discrete optimization purposes, generates a sequence of points having the same limiting distribution. (See e.g. [Aarts & Korst, 1988]). So, in this sense our class of hide-and-seek algorithms subsumes simulated annealing. Our parameter λ corresponds to the reciprocal of the temperature used in simulated annealing.

For a given value of λ, the probability that a point generated according to the density (??) lies within a ball with radius ε around a global optimum x^* (denoted by $B_\varepsilon(x^*)$) is given by

$$\Pr\{X \in B_\varepsilon(x^*)\} = \frac{\int_{B_\varepsilon(x^*)} e^{g(x;\lambda)} \, dx}{\int_S e^{g(x;\lambda)} \, dx}.$$

We are now interested in how the value of λ changes with the dimension of the problem and the precision parameter ε, if we want this probability to be at least p.

We first analyze this problem for the case

$$f(x) = -x'x.$$

The density function g then corresponds to a (truncated) normal density. Now suppose that $x^* = 0$, the global optimum on $\mathcal{R}^n$, is an interior point of S, and define $B_\varepsilon = B_\varepsilon(0)$. Then we have:

$$\Pr\{X \in B_\varepsilon\} \geq \frac{\int_{B_\varepsilon} e^{-\lambda x'x}\, dx}{\int_{\mathcal{R}^n} e^{-\lambda x'x}\, dx}$$

for ε small enough. If we use this lower bound on the probability, we are in fact assuming that the elements X_i of X are i.i.d. according to a normal distribution with mean 0 and variance $1/2\lambda$. Performing the transformation $Y_i = \sqrt{2\lambda}\, X_i$ gives:

$$\Pr\{X \in B_\varepsilon\} = \Pr\{Y \in B_{\sqrt{2\lambda}\varepsilon}\} = \Pr\{\|Y\|^2 \leq 2\lambda\varepsilon^2\}.$$

We know that the distribution of $\|Y\|^2$ is $\chi^2(n)$, and so this probability is equal to $p \in (0,1)$ if

$$2\lambda\varepsilon^2 = y(p; n)$$

or, equivalently, if

$$\lambda = \frac{y(p; n)}{2\varepsilon^2}$$

where $y(p; n)$ is chosen such that

$$\Pr\{\|Y\|^2 \leq y(p; n)\} = p.$$

Using the central limit theorem, we can for large n approximate $y(p; n)$ as follows:

$$y(p; n) \approx n + \sqrt{2n}\, z(p)$$

where $z(p)$ is chosen such that

$$\Pr\{Z \leq z(p)\} = p$$

where Z has a standard normal distribution. So, for large n, we have the following (approximate) result: If

$$\lambda \geq \frac{n + \sqrt{2n}\, z(p)}{2\varepsilon^2}$$

then

$$\Pr\{X \in B_\varepsilon\} = \Pr\{\|X - x^*\| \le \varepsilon\} \ge p.$$

So, we have $\lambda = O(n/\varepsilon^2)$.

Suppose now that, apart from the conditions mentioned above, our objective function f satisfies the following conditions:

1. $-H(x^*)$ (the negative of the Hessian matrix at the optimum x^*) is positive definite.

2. x^* is an interior point of the feasible region S.

3. ε is chosen such that f can be approximated by a quadratic function in the region $B_\varepsilon(x^*)$.

Conditions 2 implies that the gradient at x^* is zero. Using a Taylor expansion of f at x^* we get:

$$f(x) \approx f(x^*) + \frac{1}{2}(x - x^*)'H(x - x^*)$$

if x is "close" to x^*.

If furthermore the following inequality holds:

$$\int_S e^{\lambda f(x)}\, dx \le \int_S e^{\lambda(f(x^*)+\frac{1}{2}(x-x^*)'H(x-x^*))}\, dx$$

then we can generalize the abovementioned result as follows: If

$$\lambda \ge \frac{y(p;n)}{2\varepsilon^2 \theta_{min}}$$

then

$$\Pr\{X \in B_\varepsilon(x^*)\} = \Pr\{\|X - x^*\| \le \varepsilon\} \ge p$$

where $\theta_{min} > 0$ is the smallest eigenvalue of $-H(x^*)$. Using the same approximation of $y(p;n)$ as above we obtain again $\lambda = O(n/\varepsilon^2)$.

References

Aarts, E.H.L., and J.H.M. Korst. 1988. *Simulated Annealing and Boltzmann Machines.* Wiley, Chichester.

Berbee, H.C.P., C.G.E. Boender, A.H.G. Rinnooy Kan, C.L. Scheffer, R.L. Smith, and J. Telgen. 1987. Hit-and-run algorithms for the identification of nonredundant linear inequalities. *Mathematical Programming 37*, 184–207.

Boender, C.G.E., R.J. Caron, J.F. McDonald, A.H.G. Rinnooy Kan, H.E. Romeijn, R.L. Smith, J. Telgen, and A.C.F. Vorst. 1989. Shake-and-bake algorithms for generating uniform points on the boundary of bounded polyhedra. Technical Report 89–24, Department of Industrial and Operations Engineering, The University of Michigan, Ann Arbor, Michigan.

Boender, C.G.E., A.H.G. Rinnooy Kan, H.E. Romeijn, and A.C.F. Vorst. 1988. Shake-and-bake algorithms for generating uniform points on the boundary of bounded polyhedra. Technical Report 8826/A, Econometric Institute, Erasmus University Rotterdam, The Netherlands.

Boneh, A., and A. Golan. 1979. Constraints redundancy and feasible region boundedness by random feasible points generated. Paper presented at EURO III.

Dixon, L.C.W., and G.P. Szegö, eds. 1978 *Towards Global Optimization 1 and 2.* North-Holland, Amsterdam, The Netherlands.

Doob, J.L. 1953. *Stochastic Processes.* Wiley, New York.

McDonald, J.F. 1989. Monte Carolo procedures for generating points uniformly distributed over the surface of a bounded convex region. Technical Report 89–03, Department of Mathematics and Statistics, University of Windsor, Windsor, Ontario, Canada.

Pincus, M. 1968. A closed form solution of certain programming problems, *Operations Research 16*, 690–694.

Rinnooy Kan, A.H.G. and G.T. Timmer. 1984. Stochastic methods for global optimization. *American Journal of Mathematical and Management Sciences 4*, 7–39.

Rubinstein, R.Y. 1981. *Simulation and the Monte Carlo Method.* Wiley, New York.

Smith, R.L. 1980. Monte Carlo procedures for generating random feasible solutions to mathematical programs. Paper presented at the ORSA/TIMS conference, Washington, DC.

Smith, R.L. 1984. Efficient Monte Carlo procedures for generating random feasible points uniformly over bounded regions. *Operations Research 32*, 1296–1308.